CHEMICAL ENGINEERING TRANSACTIONS
VOL. 62, 2017
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet
Guest Editors: Fei Song, Haibo Wang, Fang He

ISBN 978-88-95608- 60-0; **ISSN** 2283-9216
DOI: 10.3303/CET1762141

Stoichiometry Software Services Based on Cloud Computation

Hongyu Chen

Academy of electronic information engineering, Chongqing Technology and Business Institute, Chongqing 401520, China
chenhongyu689@163.com

In order to solve the shortcomings of the existing stoichiometry software program, such as high development cost, difficulties in deployment and upgrading and poor controllability, a stoichiometry software service based on cloud computation is proposed in this paper. With software-as-a-service (SaaS) model used, the browser/server architecture is adopted to provide professional stoichiometry software services. Experiments show that the parallel cross validation framework on the platform has greatly improved with regard to the speedup ratio on quad core CPU. Therefore, CloudChem can overcome the shortcomings of traditional stoichiometry software and the software service platform based on this method can realize effective, high-speed and integral storage, analysis and digging regarding chromatography, spectroscopy, NMR, mass spectrometry and other data and minimize infrastructure cost in using the stoichiometry software and software cost.

1. Introduction

Stoichiometry which is a theory and method using statistics, mathematics, computer science and other related disciplines, is used to extract useful chemical information to the maximum from the chemical measurement data and widely used in spectroscopy, chromatography and mass spectrometry data processing (Amiri et al., 2017). With stricter requirements from the society on product quality control, lean production & safety control and continuous development of analyzer technology, more and more data is growing at an amazing rate, and this growth trend will continue to maintain in quite a long period of time (Anarado and Andreopoulos, 2016).
To this end, Cloud Chem – a cloud computing based stoichiometry software service is proposed in this paper. SaaS (Software-as-a-service) and B/S (Browser/Server) structure are used to carry out centralized storage management on data, analysis method and results. Besides, parallel computation is adopted to improve calculation speed and to achieve data and model sharing through a centralized platform (Antequera et al., 2017). The software service platform using this method can realize efficient, high-speed, integrated storage, analysis and mining of data such as spectrum, chromatography, nuclear magnetic resonance and mass spectrometry (Baktir et al., 2017).

2. Overall structure

The whole system is composed of two parts—client software and server core computing platform (Casellas et al., 2017). The client software does not provide the calculation function, but provides data compression and decompression functions and upgrades compressed data to the cloud computation center for analysis and processing, in order to reduce overhead of data transmission on the network, at the same time, analysis results can be deduced (Charalampidis et al., 2014). The client software, in addition to providing data import and export functions, should also provide a friendly interface for users to choose the analysis method, and should display analysis results through tables, 2D diagrams or 3D diagrams (Deng et al., 2015). The core part of the software is a cloud computation center, which is divided into the data node which mainly provides distributed file storage and distributed computing, and the server node which keeps metadata of mass spectrometric data and scheduling data nodes for computation (Doyle et al., 2017). The server node, the central hub of the whole system, is responsible for distributing data type conversion, mass spectrum data

Please cite this article as: Hongyu Chen, 2017, Stoichiometry software services based on cloud computation, Chemical Engineering Transactions, 62, 841-846 DOI:10.3303/CET1762141

preprocessing, quantitative analysis, qualitative analysis, data mining, generating results reports and other functions to the data node (Esposito et al., 2016). The Cloud Chem data terminal system framework is shown in Figure 1.

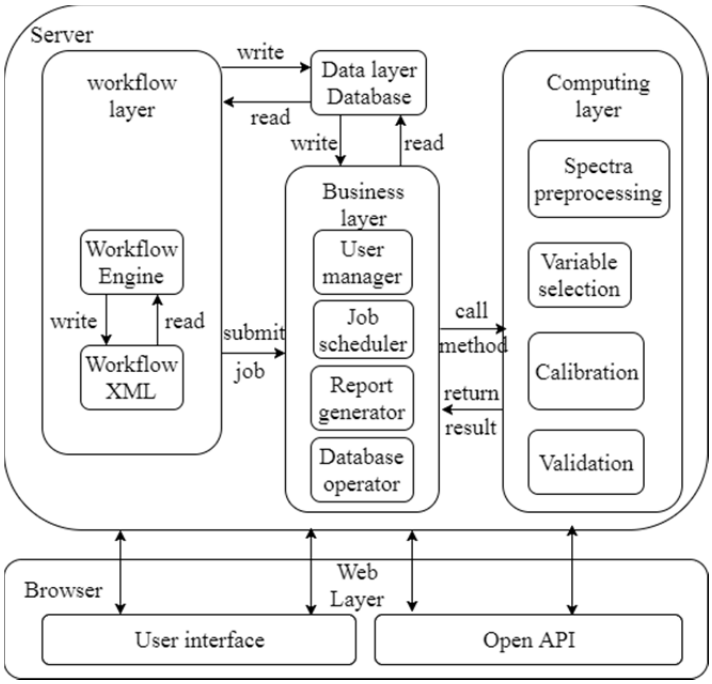

Figure 1: Cloud Chem System Framework

3. Key technology

3.1 Data storage management

The mass spectra of a single sample vary from tens to hundreds of megabytes. Usually, once the mass spectrum data files are stored on the computer, they only need to be read without data modified (Feng et al., 2016). In view of this characteristic, Cloud Chem platform will comprehensively use relational database and distributed file system to store data (Gracia-Tinedo et al., 2016). Standard data query language SQL is used by relational database to perform fast retrieval, modification and deletion of data in the database. The distributed file system using HDFS (Hadoop Distributed File System), in order to improve the data storage capacity of the system, makes it easy to expand storage space, improves the throughput of the system and also has a fast, automatic error detection and recovery function (He et al., 2017). Reliability of data storage should also be guaranteed in the wrong circumstances; besides, robustness and data integrity are also equipped, moreover, it can use Map/Reduce to realize parallel data processing (He et al., 2017).
A relational view of the experimental data table is shown in Figure 2.

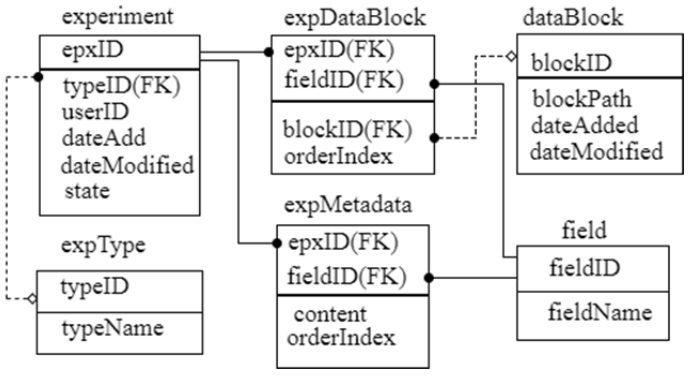

Figure 2: Relational View of Experimental Data Table

It is mainly divided into two parts: experimental description information and experimental data. In order to improve the system scalability and meet the different needs of experimental data at the same time, experimental description information is stored by this system in the expMetadata table (Huang et al, 2016). For different types of experimental data, there will be different description items that are pre-entered into the field table when configuring the system. ExpID (Experiment ID) and field ID (description item ID) are used to determine a description of the experimental data. An experiment can correspond to multiple description items (Kretsis et al., 2014), i.e. there are multiple records in the expMetadata table (Lacoste et al, 2016). The experimental data satisfies the “Write Once Read Many” file access model, so the experimental data is packaged and stored in HDFS, with data paths stored in the data Block tables, while expData Block table is used to associate experiments with data blocks (Lei et al., 2015). This scheme can retrieve the experimental description information and experimental data quickly through EPX ID.

3.2 Parallel computation

Traditional stoichiometry software runs in serial mode, which can only deal with single task and can not complete many different computing tasks in parallel, so it has limited computing speed and a difficulty in performance optimization (Li et al., 2015). Cloud Chem uses distributed computation and parallel computation to improve data processing speed and scale. It assigns different tasks submitted by different users to different servers for separate calculation, and can handle different tasks submitted by multiple users simultaneously (Mori et al., 2016). When a computational task is assigned to the server, the server uses a parallel algorithm based on multi-core systems to speed up (Pan et al., 2017).

N-fold cross-validation which is commonly used in the most typical PLS algorithm calculation model validation of the infrared spectra analysis is used as an example to briefly describe parallel stoichiometry algorithm realization method based on multi core systems (Tudoran et al., 2016). N-fold cross validation is to divide the data into N parts, where N-1 parts are used as the training set to establish the model and the remaining 1 part is subject to the newly built model to predict and calculate the prediction error (Wang et al., 2017). The process is executed in loop until each sample has been predicted only once. High speed parallelization of PLS cross validation algorithm can be implemented by using Map Reduce strategy. In general, Map Reduce divides a task on space into several independent subtasks which are calculated simultaneously by using Map operation; after <Key, Value> pairs are generated (Wang et al., 2017), the Reduce operation is used to merge the computation results of the same Key. In this paper, the parallel PLS cross validation algorithm based on Map Reduce does not require Reduce operation, and its idea is as follows:

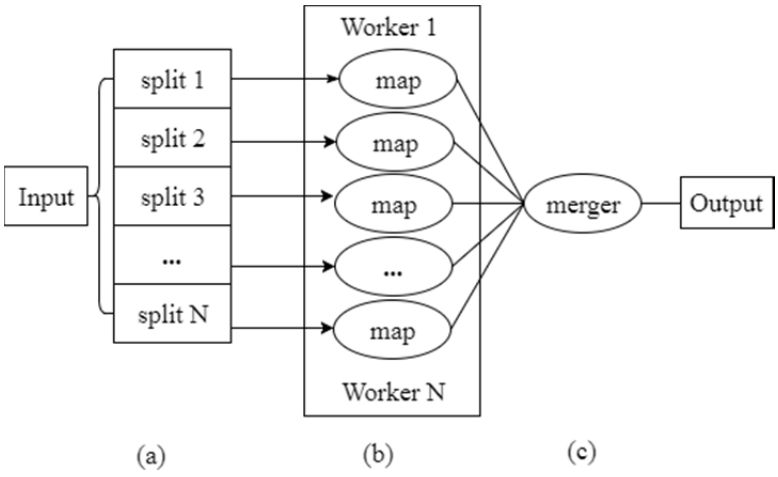

Figure 3: Parallel Framework for PLS Cross Validation

3.2.1 Split:

the sample set will be split into N parts equally (Wang et al., 2017), where sample i is a prediction set, N-1 parts constitutes a training set (Wu et al., 2017). With combination of the sample set and the training set (Xu et al., 2014), N groups of different combinations are obtained, as shown in Figure 3(a).

3.2.2 Calculation:

each combination is delivered to a Worker for individual modelling and prediction (Zeng et al., 2016), and the prediction error is calculated for which the prediction error of N models can be obtained in total, as shown in Figure 3(b).

3.2.3 Merge:
the average of prediction error completed by N Workers is calculated, and the mean value is regarded as the performance index of regression model (Zhang et al., 2017), as shown in Figure 3(c).
On the hardware platform with 4-core Intel Xeon X3430 and 4G CPU memory, with the tobacco dataset (256 x 1556) reported in the literature used (Zhang et al., 2017), the cross validation algorithm of leave-one-out method is used for tests with the number of work assumed as 1, 2, 3, 4. After repeated execution for 20 times, the average execution time is taken as real test results which are shown in Table 1.

Table 1: Speedup Efficiency of Parallel Cross Validation Algorithm

Number of worker threads	time/s	Speedup ratio	Accelerated efficiency
1	79.892	1.000	1.000
2	40.289	1.983	0.992
3	28.224	2.831	0.944
4	21.232	3.763	0.941

Speedup ratio and speedup efficiency of three SVM data sets with different size under different working threads are calculated with the speedup ratio and the speedup efficiency separately shown in Table 2 and 3.

Table 2: Speedup Ratio of Different Data Sets under Different Working Threads

	a1a	a5a	a7a
2 works	1.983039	1.979069	1.983592
3 works	2.883864	2.897889	2.894769
4 works	3.667964	3.785801	3.781785

Table 3: Speedup Efficiency of Different Data Sets under Different Working Threads

	a1a	a5a	a7a
2 works	0.991520	0.989535	0.991796
3 works	0.961288	0.965963	0.964923
4 works	0.916991	0.946450	0.945446

3.3 Open API

In order to share data and model in a standard way, REST (Representational StateTransfer) architecture, one of the salient advantages of which is that it can be implemented entirely through the HTTP protocol and it can use Cache to improve response speed is adopted by the Cloud Chem open platform. REST abstracts everything on the network as a resource, each of which corresponds to a unique resource identifier. (Zhu et al., 2012) All Open API on Cloud Chem platform can be implemented by means of POST or GET in HTTP protocol, so almost all computing languages can communicate with REST server by using HTTP protocol.

4. Conclusion

Cloud Chem-a cloud-based stoichiometry software service that employs a software-as-a-service model and uses browser/server architecture to provide professional stoichiometry software services is presented in this paper. Experiments show that the speedup ratio of the parallel cross validation framework on the 4 core CPU is greatly improved. Therefore, Cloud Chem can overcome the shortcomings of traditional stoichiometry software and the software service platform based on this method can realize effective, high-speed and integral storage, analysis and digging regarding chromatography, spectroscopy, NMR, mass spectrometry and other data and minimize infrastructure cost in using the stoichiometry software and software cost.
However, in practical application, data obtained by the same sample measured by different instruments may be different, especially in the applications of near infrared spectroscopy, this difference is more obvious. So, to share the NIR calibration model, further research is needed on the correction method of data to achieve independence of data analysis and instruments.

Reference

Amiri M., Sobhani A., Osman H.A., Shirmohammadi S., 2017, SDN-Enabled Game-Aware Routing for Cloud Gaming Datacenter Network, IEEE Access, 5, 18633-18645, DOI: 10.1109/ACCESS.2017.2752643

Anarado I., Andreopoulos Y., 2016, Core Failure Mitigation in Integer Sum-of-Product Computations on Cloud Computing Systems, IEEE Transactions on Multimedia, 18, 789-801, DOI: 10.1109/TMM.2016.2532603

Antequera R.B., Calyam P., Debroy S., Cui L., Seetharam S., Dickinson M., Joshi T., Xu D., Beyene T., 2017, ADON: Application-Driven Overlay Network-as-a-Service for Data-Intensive Science, IEEE Transactions on Cloud Computing, 1, DOI: 10.1109/TCC.2015.2511753

Baktir A.C., Ozgovde A., Ersoy C., 2017, How Can Edge Computing Benefit from Software-Defined Networking: A Survey, Use Cases, and Future Directions, IEEE Communications Surveys Tutorials, 19, 2359-2391, DOI: 10.1109/COMST.2017.2717482

Casellas R., Vilalta R., Martínez R., Muñoz R., 2017, Highly available SDN control of flexi-grid networks with network function virtualization-enabled replication, IEEE/OSA Journal of Optical Communications and Networking, 9, A207-A215, DOI: 10.1364/JOCN.9.00A207

Charalampidis P., Albano M., Griffiths H., Haddad A., Waters R.T., 2014, Silicone rubber insulators for polluted environments part 1: enhanced artificial pollution tests, IEEE Transactions on Dielectrics and Electrical Insulation, 21, 740-748, DOI: 10.1109/TDEI.2013.004015

Deng S., Huang L., Taheri J., Zomaya A.Y., 2015, Computation Offloading for Service Workflow in Mobile Cloud Computing, IEEE Transactions on Parallel and Distributed Systems, 26, 3317-3329, DOI: 10.1109/TPDS.2014.2381640

Doyle J., Giotsas V., Anam M.A., Andreopoulos Y., 2017, Dithen: A Computation-as-a-Service Cloud Platform For Large-Scale Multimedia Processing, IEEE Transactions on Cloud Computing, 1, DOI: 10.1109/TCC.2016.2617363

Esposito C., Castiglione A., Choo K.K.R., 2016, Challenges in Delivering Software in the Cloud as Microservices, IEEE Cloud Computing, 3, 10-14, DOI: 10.1109/MCC.2016.105

Feng B., Ma X., Guo C., Shi H., Fu Z., Qiu T., 2016, An Efficient Protocol with Bidirectional Verification for Storage Security in Cloud Computing, IEEE Access, 4, 7899-7911, DOI: 10.1109/ACCESS.2016.2621005

Gracia-Tinedo R., García-López P., Sánchez-Artigas M., Sampé J., Moatti Y., Rom E., Naor D., Nou R., Cortés T., 2016, IOStack: Software-Defined Object Storage, IEEE Internet Computing, 20, 10-18, DOI: 10.1109/MIC.2016.46

He D., Kumar N., Khan M.K., Wang L., Shen J., 2017, Efficient Privacy-Aware Authentication Scheme for Mobile Cloud Computing Services, IEEE Systems Journal, 1-11, DOI: 10.1109/JSYST.2016.2633809

He S., Wang Y., Sun X.H., Xu C., 2017, Using MinMax-Memory Claims to Improve In-Memory Workflow Computations in the Cloud, IEEE Transactions on Parallel and Distributed Systems, 28, 1202-1214, DOI: 10.1109/TPDS.2016.2614294

Huang G., z.Liu X., Lu X., Ma Y., Zhang Y., Xiong Y., 2016, Programming Situational Mobile Web Applications with Cloud-Mobile Convergence: An Internetware-Oriented Approach, IEEE Transactions on Services Computing, 1, DOI: 10.1109/TSC.2016.2587260

Kretsis A., Christodoulopoulos K., Kokkinos P., Varvarigos E., 2014, Planning and operating flexible optical networks: Algorithmic issues and tools, IEEE Communications Magazine, 52, 61-69, DOI: 10.1109/MCOM.2014.6710065

Lacoste M., Miettinen M., Neves N., Ramos F.M.V., Vukolic M., Charmet F., Yaich R., Oborzynski K., Vernekar G., Sousa P., 2016, User-Centric Security and Dependability in the Clouds-of-Clouds, IEEE Cloud Computing, 3, 64-75, DOI: 10.1109/MCC.2016.110

Lei X., Liao X., Huang T., Li H., 2015, Cloud Computing Service: The Caseof Large Matrix Determinant Computation, IEEE Transactions on Services Computing, 8,688-700, DOI: 10.1109/TSC.2014.2331694

Li J., Tan X., Chen X., Wong D.S., Xhafa F., 2015, OPoR: Enabling Proof of Retrievability in Cloud Computing with Resource-Constrained Devices, IEEE Transactions on Cloud Computing, 3, 195-205, DOI: 10.1109/TCC.2014.2366148

Mori S., Wu D., Ceritoglu C., Li Y., Kolasny A., Vaillant M.A., Faria A.V., Oishi K., Miller M.I., 2016, MRICloud: Delivering High-Throughput MRI Neuroinformatics as Cloud-Based Software as a Service, Computing in Science Engineering, 18, 21-35, DOI: 10.1109/MCSE.2016.93

Pan W., Zheng F., Zhao Y., Zhu W.T., Jing J., 2017, An Efficient Elliptic Curve Cryptography Signature Server with GPU Acceleration, IEEE Transactions on Information Forensics and Security, 12, 111-122, DOI: 10.1109/TIFS.2016.2603974

Tudoran R., Costan A., Antoniu G., 2016, OverFlow: Multi-Site Aware Big Data Management for Scientific Workflows on Clouds, IEEE Transactions on Cloud Computing, 4, 76-89, DOI: 10.1109/TCC.2015.2440254

Wang H., Li Y., Zhang Y., Jin D., 2017, Virtual Machine Migration Planning in Software-Defined Networks, IEEE Transactions on Cloud Computing, 1, DOI: 10.1109/TCC.2017.2710193

Wang K., Yang K., Chen H.H., Zhang L., 2017, Computation Diversity in Emerging Networking Paradigms, IEEE Wireless Communications, 24, 88-94, DOI: 10.1109/MWC.2017.1600161WC

Wang W., Jiang Y., Wu W., 2017, Multiagent-Based Resource Allocation for Energy Minimization in Cloud Computing Systems, and Cybernetics: Systems IEEE Transactions on Systems, Man, 47, 205-220, DOI: 10.1109/TSMC.2016.2523910

Wu T., Dou W., Hu C., Chen J., 2017, Service Mining for Trusted Service Composition in Cross-Cloud Environment, IEEE Systems Journal, 11, 283-294, DOI: 10.1109/JSYST.2014.2361841

Xu Z., Wang C., Ren K., Wang L., Zhang B., 2014, Proof-Carrying Cloud Computation: The Case of Convex Optimization, IEEE Transactions on Information Forensics and Security, 9, 1790-1803, DOI: 10.1109/TIFS.2014.2352457

Zeng D., Gu L., Guo S., Cheng Z., Yu S., 2016, Joint Optimization of Task Scheduling and Image Placement in Fog Computing Supported Software-Defined Embedded System, IEEE Transactions on Computers, 65, 3702-3712, DOI: 10.1109/TC.2016.2536019

Zhang L., Jung T., Liu K., Li X.Y., Ding X., Gu J., Liu Y., 2017, PIC: Enable Large-Scale Privacy Preserving Content-Based Image Search on Cloud, IEEE Transactions on Parallel and Distributed Systems, 28, 3258-3271, DOI: 10.1109/TPDS.2017.2712148

Zhang Y., Xu C., Liang X., Li H., Mu Y., Zhang X., 2017, Efficient Public Verification of Data Integrity for Cloud Storage Systems from Indistinguishability Obfuscation, IEEE Transactions on Information Forensics and Security, 12, 676-688, DOI: 10.1109/TIFS.2016.2631951

Zhu W., Lu W., Zhang X., Cai Z., Liu H., Peng L., Li H., Han Y., Fen W., 2012, Nano-hydroxyapatite/fibrin glue/recombinant human osteogenic protein-1 artificial bone for repair of bone defect in an animal model, IET Micro Nano Letters, 7, 467-471, DOI: 10.1049/mnl.2012.0090